Cross-correlation in face discrimination

Simpson, William A.; Loffler, Gunter; Tucha, Lara

Published in:
Vision Research

DOI:
10.1016/j.visres.2012.10.014

Vision Research 76 (2013) 60–67

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Cross-correlation in face discrimination

William A. Simpson [a,*], Gunter Loffler [b], Lara Tucha [c]

[a] School of Psychology, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK
[b] Department of Vision Sciences, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, UK
[c] Department of Clinical and Developmental Neuropsychology, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 14 December 2011
Received in revised form 8 October 2012
Available online 1 November 2012

Keywords:
Face discrimination
Cross-correlation
Face perception

ABSTRACT

An extensive body of literature suggests that face perception depends critically upon specialised face processing mechanisms. Although it seems clear that specialised face processing is required to explain face recognition, face discrimination is a simpler task that could possibly be solved with a general pattern discrimination mechanism. Observers were presented with face images that were either identical or not and judged whether they were the same or different. Face discrimination performance was well described by the point-by-point cross-correlation between the face images, which is a simple mechanism of the type used for discriminating patterns such as gratings. This result held for male and female faces viewed frontally or in profile. Results for inverted and contrast-reversed faces were also well described by cross-correlation, with observers having lowered efficiency relative to normal faces.

1. Introduction

Humans are very good at telling faces apart, and much research has been done in an effort to find the mechanisms underlying face perception. There is ample evidence from cell recording (Desimone, 1991), lesions (Barton et al., 2002), and fMRI (Kanwisher, McDermott, & Chun, 1997; Loffler, Yourganov et al., 2005) that extrastriate areas such as Fusiform Face Area (FFA) contain specialised mechanisms that are critically important for face recognition. However, the existence of a high-level face processing centre does not necessarily mean that it is required for all perceptual tasks related to faces or that its computations always explain observers' performance. For example, if, in a detection task, an observer is asked to report whether a face or a blank field with the same mean luminance is presented, performance may be determined by the sensitivity of neurons downstream from FFA (e.g. primary visual cortex, V1), rather than by the computation of a network specialised for faces. Face detection and recognition may be considered as lying on opposite ends of the spectrum of complexity for face computations. Face discrimination – saying whether two faces are the same or different – is somewhere in between. Unlike recognition, discrimination does not entail comparison with a set of faces stored in memory. Discrimination inherently measures perceptual rather than memory processing, and yet is a more sophisticated ability than detection. In this paper we will investigate whether face discrimination might be explained using a very simple-minded cross-correlation scheme of the sort used for discrimination of basic non-face patterns such as sine wave gratings (Burgess & Ghandeharian, 1984).

In visual detection and discrimination, ample evidence exists for the idea that humans act as template matchers for simple patterns (Eckstein & Ahumada, 2002). Indeed, for these tasks the very best performance possible, that of an ideal observer, is achieved by a template matching strategy (Simpson, Falkenberg, & Manahilov, 2003; van Trees, 1968; Whalen, 1971). The performance of a template-matcher is limited by the cross-correlation of the patterns to be discriminated. In a simple discrimination experiment, the observer needs to decide if a presented stimulus corresponds to signal s_0 or s_1. These two signals are known exactly by the ideal observer. The ideal observer will cross-correlate the stimulus with s_0 and with s_1, and will say "s_0" if it produces the greater cross-correlation, and "s_1" otherwise. The situation is more complicated in a same–different experiment where random pairings of a large number of unknown stimuli are presented. We are not aware of an ideal observer solution for this problem. A sensible strategy is for the observer to compute a similarity measure between the two images, and to say "different" if the similarity is below some cut-off and "same" otherwise. The similarity could be computed as the cross-correlation or the sum of the squared intensity differences (which amounts to the same thing as cross-correlation) or by correlations over filter banks (Biederman & Kalocsai, 1997). If humans can use such a simple method in discriminating faces, we would expect their performance to be strongly predicted by the cross-correlation between face images. In the experiments to be reported

* Corresponding author.
E-mail addresses: william.simpson@plymouth.ac.uk (W.A. Simpson), G.Loffler@gcu.ac.uk (G. Loffler), l.i.tucha@rug.nl (L. Tucha).

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2012.10.014

here, we measured face discrimination performance using male, female, frontal and profile faces, and compared human performance to that of a cross-correlator.

2. Experiment 1

In Experiment 1 we presented observers with male and female faces in frontal and profile views, and they performed a same/different discrimination task. If face discrimination can be explained by a simple cross-correlation algorithm, the probability of a correct judgement should be well described by a generalized linear model using the cross-correlation between each pair of faces as the predictor.

Two types of face samples were used. In one condition, the faces were drawn from a larger set. In another condition, the faces were created as a morph series, which allowed a denser sampling of face cross-correlation values in a region that would be sparse otherwise.

2.1. Method

2.1.1. Participants

Psychology undergraduates from the University of Plymouth participated in exchange for partial course credit. There were 57 participants for the random face stimuli and 69 participants for the morph series face stimuli.

2.1.2. Stimuli

We used frontal and profile views of male and female faces from the Max-Planck Institute for Biological Cybernetics database (Troje & Bülthoff, 1996). Each face was contained in a 256 × 256 pixel greyscale (8 bits per pixel) image (see Fig. 1). All faces were centred in the square image region, with some variation in the height and width of each face. Images were cropped to exclude head shape and hair. Faces had neutral expressions and all observers were unfamiliar with the individuals shown on the images. For the linear interpolation (morph) sequence, two female frontal faces were chosen as the end-points. These were used to construct a morph sequence consisting of 10 faces in total, using Morpher 3.1. At the viewing distance of 80 cm (from a chin- and forehead-rest), the 256 × 256 pixel region subtended a visual angle of 8.4°, though the faces themselves were roughly half this size. The faces were presented side by side, and the centres of the faces were separated by 4.2°. The fixation mark in the centre of the screen was 0.02° square. The stimuli were displayed on a Viewsonic CRT monitor at a refresh rate of 100 Hz. The mean luminance of the display was 64 cd/m^2, and the Weber contrast was 95%.

Fig. 1. A typical pair of stimulus faces. The example shows two female faces in frontal view. Stimuli included all combinations of gender (male and female) and view (frontal and profile) but the gender and view were always matched between the two faces to be discriminated and different combinations were run in separate blocks. The observers' task was to decide if the two presented faces were the same or not.

2.1.3. Procedure

The observer viewed the monitor from a chin- and forehead-rest. Each trial consisted of the presentation of the fixation mark in the centre of the screen for 600 ms, followed by a 100 ms presentation of a pair of faces. Each block of 180 trials contained faces with a particular combination of gender and view. A fixed set of 10 faces was used in each condition (male frontal, female frontal, male profile, female profile, and morph series). All possible pairings of the faces were presented twice in random order. Equal numbers of same and different face pairs were presented. The four combinations of gender (male or female) and view (frontal or profile) were presented in blocked random order (all faces within a block shared the same gender and view). For example, the observer might first receive a 180 trial run of male frontal faces, followed by a run of female profile faces, etc. The morph series was presented to a different set of observers. The observer's task was to indicate if the two faces were the same or not, by pressing one of two buttons.

Short presentation times and a fixation mark were used in order to obviate eye movements. It is known that fine discrimination between faces can be performed with short presentation times (Lehky, 2000; Loffler, Gordon, et al., 2005).

2.2. Results and discussion

For an ideal observer, when the signals being discriminated are known, the discrimination performance is determined by the pixel-by-pixel cross-correlation between the images being compared. Although in our experiments the signals being discriminated were not known by the observers, a reasonable strategy is to cross-correlate the two signals, say "different" if the cross-correlation value is low, and "same" otherwise. Fig. 2 shows the proportion of "different" judgements for each pair of faces plotted as a function of the normalised cross-correlation between them (Lewis, 1995). All but one point in Fig. 2 is based on 114 judgements (57 observers, each face pair seen twice) for each face pair. The point for the identical face pairs (cross-correlation = 1.0) is based on 5130 judgements. There are 46 points: 45 points for the different face

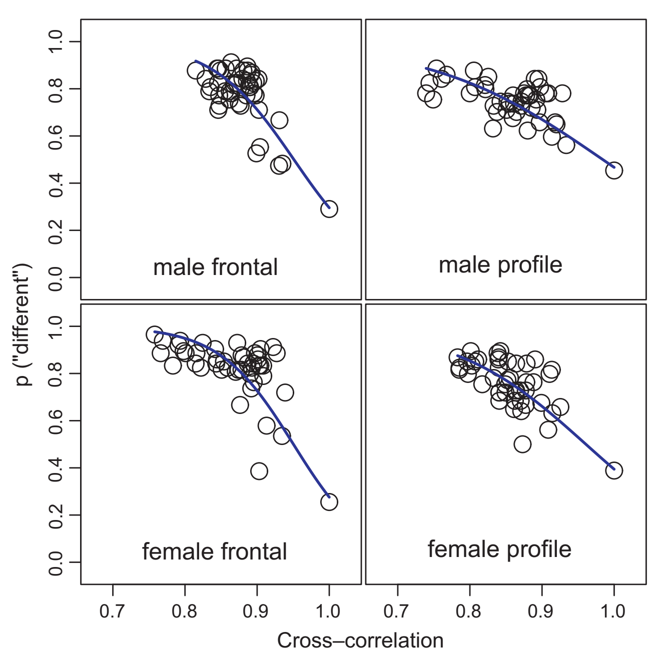

Fig. 2. Proportion "different" judgements as a function of the normalised cross-correlation between pairs of faces in a same/different task. Each point represents a single face pair. Cumulative logistic curves, as fit by generalized linear model, are shown.

pairs having unique cross-correlations, and one point for the identical face pairs (cross-correlation = 1.0). As can be seen in Fig. 2, p("different") is closely tied to the cross-correlation between the faces being compared.

Before proceeding to an analysis of the data in Fig. 2, let us consider the performance of a model observer. On each trial, the observer is presented with a pair of faces that has some value of cross-correlation. An observer who could perfectly discriminate the stimuli would respond using the rule: if the cross-correlation equals 1, say "same", else say "different". This observer's p('different') vs cross-correlation results would be described by a step function. The results shown in Fig. 2 are consistent with the observer adding internal noise to the delivered cross-correlation. According to this idea, internal noise is being added to each image (many results in simple pattern detection and discrimination support this idea; see Burgess, 1990), and thus the resulting cross-correlation is noisy as well. We assume that the noise is logistic with mean 0 and spread s. The observer will say "same" when the decision variable (cross-correlation + noise) is above some criterion, and "different" otherwise. According to this model, p("different") plotted as a function of cross-correlation will be a cumulative logistic whose mean is the criterion and whose spread is s. Generalized linear models using binomial family and logistic link produce a fit of the form

$$p(\text{“different”}) = \text{logit}(\text{intercept} + \text{slope} \cdot \text{cross-correlation}).$$

The spread of the noise is |1/slope|. The noise is what limits real observer performance, and so we can evaluate relative performance in terms of the fitted GLM slope, with a steep slope indicating good performance and low noise.

Real observers are inefficient compared to an ideal observer (Tanner & Birdsall, 1958), and internal noise is a factor in this. For the purpose of making the discussion more concrete, let us consider the case of discrimination of two known equal energy signals s_0 and s_1 (which is unlike our experimental situation where the signals are unknown). For this case the performance of the ideal observer is

$$d' = \sqrt{\frac{2E(1-\rho)}{\sigma_e^2}} \tag{1}$$

where d' is the usual signal detection theory measure of discriminability, E is the energy of each signal, ρ is the cross-correlation between the signals, and σ_e^2 is the external noise variance (Simpson, Falkenberg, & Manahilov, 2003). The efficiency η (Tanner & Birdsall, 1958) of a real observer is

$$\eta = \frac{d'^2_{\text{real}}}{d'^2_{\text{ideal}}} \tag{2}$$

where d'^2_{real} and d'^2_{ideal} are the performance measured from real observers and predicted from ideal observers for the presented stimuli. Burgess proposed that the lower than unity efficiency found for real observers was due to two factors: internal noise σ_i^2 and sampling efficiency k (Burgess, 1990; Burgess et al., 1981). The resulting real observer model is

$$d' = \sqrt{\frac{k2E(1-\rho)}{\sigma_e^2+\sigma_i^2}}. \tag{3}$$

In this framework, the real observer only uses a fraction of the presented stimulus energy, and he adds extra noise to the stimuli. Substituting (1) and (3) into (2), we have

$$\eta = \frac{2E(1-\rho)}{\sigma_e^2} \Big/ \frac{k2E(1-\rho)}{\sigma_e^2+\sigma_i^2} = \frac{k}{1+\frac{\sigma_i^2}{\sigma_e^2}}. \tag{4}$$

Eq. (4) makes explicit the relation between the *overall efficiency* η, the *sampling efficiency* k, and the internal noise σ_i^2. A given level of overall efficiency will be less than unity due to sampling efficiency < 1 or internal noise > 0, or both.

Now, let us relate this treatment to the slopes of our psychometric functions. The measure d' is defined by

$$d' = Q(p(\text{hit})) - Q(p(\text{false alarm})) \tag{5}$$

where $Q()$ is the normal quantile function, p(hit) in our case represents the judgement "different" when the two stimuli are different ($\rho < 1$), and p(false alarm) represents the judgement "different" when the two stimuli are identical ($\rho = 1$). The inverse of $Q()$ is the distribution function CDF(), giving

$$p(\text{“different”}) \mid \rho < 1 = \text{CDF}(d' + Q(p(\text{“different”}) \mid \rho = 1)) \tag{6}$$

where d' is given by Eq. (3). The proportion of "different" judgements as a function of the cross-correlation between the stimuli will be an ogive with a sideways shift (intercept) dependent upon the observer's criterion, and a slope dependent upon the internal noise variance and sampling efficiency. The above analysis for the case of signals known exactly was presented in order to make the discussion concrete. However, this joint action of the sampling efficiency and internal noise on the psychometric function slope is a general effect and not limited to the signal known exactly case.

The psychometric functions in Fig. 2 shows that the observers act as though noise was present in the stimuli. Without it, they would be step functions, because there was no external noise in the stimuli. What such noise actually consists in (e.g. fluctuations in attention or criterion) and where it comes from is unknown. If we assume the non-ideal observer framework proposed by Burgess, changes in slope between experimental conditions could be due to changes in internal noise, changes in efficiency or both. We define the *effective noise* σ^2_{eff} as

$$\sigma^2_{\text{eff}} = (\sigma_e^2 + \sigma_i^2)/k = 1/\text{slope}^2$$

The effective noise that we observe in the psychometric function is determined by some non-zero amount of internal noise (because $\sigma_e^2 = 0$) whose effect is scaled by the sampling efficiency. For example, if the internal noise is fixed, lowering the sampling efficiency produces an effective noise that is larger.

Generalized linear models using binomial family and logistic link were fit to the same/different data for each condition (combination of gender and view). The GLM fits were to the individual Bernoulli same/different points resulting from each trial (i.e. not to binomial proportions). Table 1 shows the statistics resulting from the fits. The χ^2 values and associated p-values in the table are for likelihood ratio tests each comparing a model which included cross-correlation as a predictor to an intercept-only (null) model. The likelihood ratio test indicates that the cross-correlation was a strong predictor of same/different judgements for all conditions ($p < 1\text{e}{-16}$). The strength of the relationship between cross-correlation and p("different") can be assessed by concordance ratio or c-index (Harrell et al., 1982) and by the log odds ratio. We shall consider these in turn.

The c-index measures the probability of concordance between observed responses and those predicted from the fitted model. The c-index varies from 0.5, for random performance, to 1.0 for perfect prediction. Since the observed responses are 0 s and 1 s, a c-index of 1.0 can only be obtained when the predicted responses are 0 s and 1 s. Therefore very high c-index values will never be seen when a typical psychometric function is measured. The observed c-index is substantially above 0.5 in all cases (near 0.75), confirming the conclusion from the likelihood ratio tests and from visual inspection of Fig. 2 that cross-correlation is a good predictor of same/different judgements. One might wonder what size of c-index

Table 1
Statistics for logistic regression of p("different") as a function of cross-correlation between the faces.

Condition	Intercept [95% CI]	Slope [95% CI]	χ^2	p	Concordance ratio [95% CI]
Male frontal	16.84 [16.14, 17.54]	−17.71 [−18.44, −16.97]	2748.05	<1e−16	0.76 [0.757, 0.771]
Male profile	8.26 [7.73, 8.80]	−8.40 [−8.96, −7.83]	970.96	<1e−16	0.67 [0.658, 0.675]
Female frontal	18.47 [17.73, 19.21]	−19.44 [−20.21, −18.66]	3548.48	<1e−16	0.80 [0.797, 0.810]
Female profile	10.53 [9.99, 11.08]	−10.97 [−11.55, −10.39]	1578.16	<1e−16	0.71 [0.699, 0.715]
Morph	34.57 [33.02, 36.11]	−35.44 [−37.02, −33.87]	3545.87	<1e−16	0.78 [0.770, 0.783]

should be observed if the logistic model using cross-correlation as a predictor was actually correct, and only the Bernoulli random variable nature of the responses influenced the c-index's size. This question was answered using simulations. Assuming that the true function underlying the responses was as given by the intercepts and slopes in Table 1, the c-index for male frontal should be in the 95% CI of 0.76–0.78 (observed was 0.76), for male profile 0.66–0.67 (observed was 0.67), for female frontal 0.79–0.81 (observed was 0.80) and for female profile 0.70–0.72 (observed was 0.71). The observed c-indexes were of the size expected if the data truly were generated by a logistic model having cross-correlation as the predictor.

The log odds ratio is another way to assess the strength of the relationship between the face cross-correlation and p("different"). The log odds ratio is equal to the slope in logistic regression, where it is the primary measure of effect size. The odds ratio is the change in the odds produced as the predictor increases by one unit. Take for example the male frontal data, where the slope is −17.71. This means that the odds are exp(17.71) or 49,130,963 times higher for saying "different" when the cross-correlation between the faces is 0.0 than when it is 1.0. This is a huge effect size, considering that an odds ratio of 1 (slope of 0) signifies no effect. The smallest odds ratio in Table 1 is 4447.

In fitting the logistic regressions both slopes and intercepts were obtained, and in all cases both were significantly different from zero at p < 1e−16. For male frontal faces, the fitted parameters were: intercept = 16.8, slope = −17.7. For male profile, intercept = 8.3, slope = −8.4. For female frontal, intercept = 18.5, slope = −19.4. For female profile, intercept = 10.5, slope = −11.0. As mentioned earlier, it is the slopes which are important, because the reciprocal of the slope indicates the size of the effective noise σ^2_{eff} (the spread parameter, s, of the logistic distribution; SD = $\pi s/\sqrt{3}$). It is clear from Fig. 2 that the slopes are steeper for frontal than for profile views (95% CIs are given in Table 1). Thus, the effective noise is less when the observer is computing the cross-correlation between frontal views of faces compared to profile views. Changes in the effective σ^2_{eff} noise could be caused by less internal noise being added or by increased sampling efficiency with frontal views. The improvement in performance makes sense when we consider that humans have most practice and expertise with frontal views. Frontal views are used for verification of personal identity (passports, driver's licences) for this reason.

The performance of human observers shown in Fig. 2 seems to be well fit using the cross-correlation between the stimuli as a predictor. The quality of the model fit can be further assessed by examining the deviance residuals associated with the GLM fits (see Fig. 3). If the data are adequately described by a logistic regression using cross-correlation as a predictor, we would expect the residuals to have no obvious pattern and to fall in the range ±3. The scatter of the data points about the fitted logistic regression curves is as expected from binomial random variables. If there had been less or more scatter than we observed, or if there had been obvious patterns in the residuals, there would have been cause for concern.

Although the data in Fig. 2 are fitted well by using the cross-correlation as a predictor of performance, most face pairs had

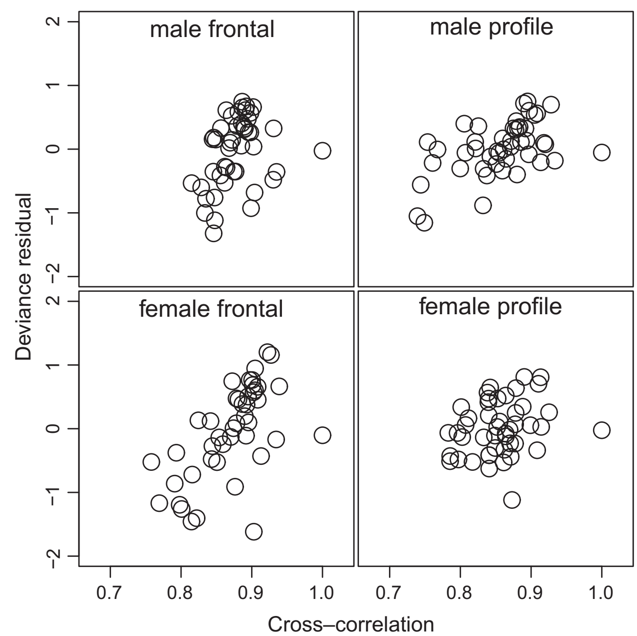

Fig. 3. Deviance residual plots for the data and GLM fits shown in Fig. 2. If the data are adequately described by a logistic regression using cross-correlation as a predictor, we would expect the residuals to have no obvious pattern and to fall in the range ±3. The model fits appear adequate.

cross-correlations of roughly 0.9 or less. This is an inevitable consequence of pairing photographs of different individuals' faces that were not systematically generated to have certain correlations. The gap in the cross-correlations between 0.9 and 1.0 was filled by using female frontal faces from a 10 face morph series (Fig. 3). Each point in the top panel of Fig. 4 shows the proportion of "different" responses out of 138 trials (69 observers, each pair seen twice), except for the point for identical faces (cross-correlation = 1.0), which is based on 6210 responses.

The results of the GLM fit are shown in the bottom row of Table 1. Both of the parameters of the logistic fit for p("different") as a function of cross-correlation, the intercept (34.57) and the slope (−35.44), were significantly different from zero at p < 1e−16. The likelihood ratio test of the model which included the cross-correlation as a predictor vs the intercept-only model gave a χ^2 value of 3545.87, p < 1e−16. The concordance ratio for predicted and obtained responses gave a value of 0.78 (95% CI: [0.770, 0.783]). Through simulations it was found that the expected c-index when the model is assumed to be correct would be in the range of 0.75–0.77. Thus the observed c-index is in the range expected if Bernoulli response variability is the only cause of lack of fit. Inspection of the top panel in Fig. 3 shows that for faces pairs having low cross-correlations, the obtained responses fall below the fitted curve. The asymptotic performance for these observers is less than 1.0. This point is made clear by the deviance residual plot in the bottom panel. Note the progressively larger negative residuals for cross-correlations less than about 0.9, which indicate

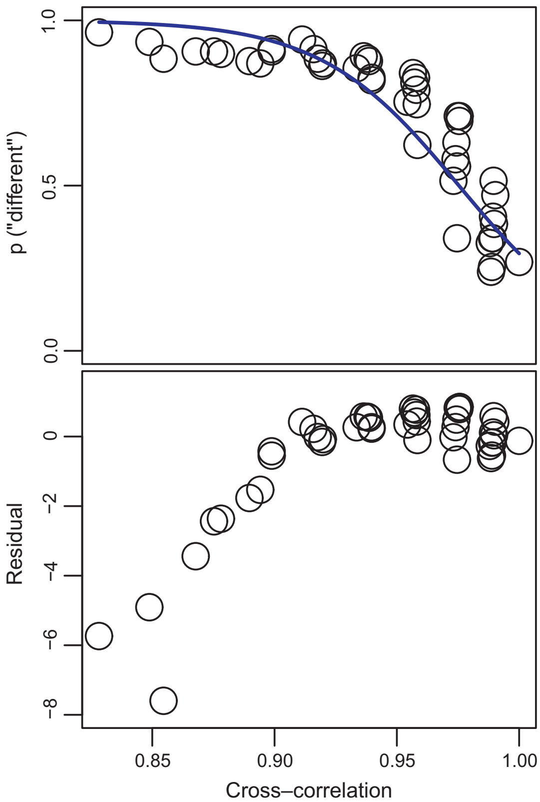

Fig. 4. Results for a same/different task in judging pairs of female frontal faces from a linear morph sequence. Top, proportion "different" judgements as a function of the cross-correlation between the pairs of faces, with fitted cumulative logistic function. Bottom, the deviance residuals show lack of fit for cross-correlation values less than 0.9. This is due to the asymptotic *p*("different") being less than 1.0 (see top plot).

that *p*("different") stays fixed when it should be increasing towards 1.0. Cross-correlation is still a good predictor of performance, however the asymptotic performance of the observers is less than perfect.

The results just discussed show that cross-correlation is a good predictor of same/different judgements. We can interpret this in more than one way. One interpretation is that the cross-correlation between the two faces is a stimulus property that serves well as a predictor of performance. From a system identification point of view, if we present pairs of faces to be discriminated to a black box (observer), the output of that black box can be predicted using cross-correlation. Another interpretation is that the visual system actually uses a more sophisticated strategy than cross-correlation, but that cross-correlation is a good descriptor of performance nonetheless. A final interpretation is that the visual system actually computes the cross-correlation between the faces. We do not favour one interpretation over any other.

To summarise Experiment 1, performance in a same/different face discrimination task was well described by the cross-correlation between the stimuli. Given a simple task, cross-correlation is sufficient to explain performance. This finding is in agreement with the neuropsychological results of Benton and Van Allen (1972), whose patient could discriminate but not recognise faces. A traditional approach to psychophysics is to regard the observer as a black box, and to attempt to characterise mathematically the function mapping stimuli onto responses. Cross-correlation is satisfactory for mapping stimuli onto responses in our experiments.

Any visual discrimination will be limited by noise. That is why any psychometric function is an ogive rather than a step function, even for noiseless stimuli. Cross-correlation is a good predictor of face discrimination performance, but the source and nature of the internal noise are also things that need explanation. We have suggested that each of the two face representations has noise added to it, and therefore the cross-correlation of the two noisy representations will be noisy as well. The question then arises: What causes the effective noise level to vary depending on whether the faces are viewed in profile or frontally? It is well known that the visual system is better at discrimination of the orientation of gratings that are near horizontal and vertical (Appelle, 1972). One possible reason for this is that the human visual system in daily life predominantly receives these orientations as stimulation. Through perceptual learning, the visual system reduces internal noise for these orientations (Dosher & Lu, 1998). Similarly, in our daily interactions with other people, we predominantly receive frontal views. The effective noise (σ^2_{eff}) is affected by internal noise (σ^2_i) and also by sampling efficiency (k). It is also possible that perceptual learning improves efficiency in addition to reducing σ^2_i.

A key difference of our stimuli from those of Wilbraham et al. (2008), who could not predict face discrimination from low-level face image properties, is that the faces truly were identical or not. Wilbraham et al. used face pairs which were never identical; two "same" faces would have different expressions or illumination. This approach reduces the possibility that a low-level computation, based on image intensities, can predict human performance. Their experimental stimuli should therefore *not* be discriminable with a simple cross-correlation mechanism. Arguably these stimuli were chosen specifically to demonstrate that under general face recognition conditions (i.e. in the presence of lighting and expression variations), complex mechanisms are required to explain performance.

3. Experiment 2

Experiment 1 showed that face discrimination performance was predicted well by the cross-correlation between the faces. One signature of face processing is its disproportionate disruption by inversion (Sekuler et al., 2004; Yin, 1969; Yovel & Kanwisher, 2005). Processing of contrast reversed (negative) faces is also impaired (Galper, 1970; Gaspar, Bennett, & Sekuler, 2008). In both cases, processing of faces is more impaired than is processing of non-face stimuli. In Experiment 2, we presented inverted and negative faces. Our reasoning is that if the discrimination task does not engage normal face processing, then we should not find impaired performance for inverted and negative faces relative to normal (i.e. upright and positive contrast) faces. We have seen in Experiment 1 that the level of effective noise is larger for discrimination of profile faces than for frontal faces. This can be attributed to an introduction of more noise or to lowered sampling efficiency during the computation of cross-correlation for profile faces. We expect the same to be true for inverted and negative faces relative to normal faces.

3.1. Method

3.1.1. Participants

Psychology undergraduates from the University of Plymouth participated in exchange for partial course credit. There were 49 participants.

3.1.2. Stimuli

Female frontal faces from the same database as in Experiment 1 were used. Each trial used a random face pair (same or different with equal probability) sampled from the 100 face database. Each observer saw a different random sampling of faces.

3.1.3. Procedure

The procedure was the same as used in Experiment 1. Participants viewed pairs of faces, and judged if they were the same or different. The faces were upright, inverted, or negative (reversed contrast) within separate blocks of trials. The order of the blocks was random. Each block contained 110 trials, with equal numbers of same and different face pairs.

3.2. Results and discussion

In Experiment 1 we found that p("different") judgements were well predicted by the cross-correlation between the faces. Now we examine how the psychometric functions are affected by face inversion and contrast reversal. We saw in Experiment 1 that observers' performance was consistent with a noisy cross-correlation, where internal noise is added during the cross-correlation process, and the strength of the effective noise σ^2_{eff} (the amount of noise equivalent to the joint effect of internal noise and sampling efficiency) is indicated by the reciprocal of the squared slope of the psychometric function. From the face literature it is known that face inversion and contrast reversal impair face recognition. This leads us to expect an increase in the effective noise level exhibited when observers discriminate inverted and contrast reversed faces.

The data are shown in Fig. 5. Each observer on each run saw a different random sampling of face pairs. There were 100 faces in the pool of faces, and there are 4950 possible face pairings with this pool. For each trial we presented a pair of faces, whose cross-correlation we computed, and we measured a same/different (0/1) response. The data analysis fit a generalized linear model with binomial family and logistic link, where each y was a Bernoulli same/different (0/1) response and each x was the normalised cross-correlation between the face pair presented on a given trial. A plot of these binary points would not be informative. Therefore Fig. 5 shows proportions of "different" responses falling in each of 31 cross-correlation bins. The bin widths were adjusted such that each bin contained 90 observations. This binning was done purely for visualisation purposes – as mentioned, the GLM fitting was done on the unbinned 0/1 responses.

As can be seen in Fig. 5, the probability of judging a pair of faces to be different was well predicted by the cross-correlation between the faces, whether these faces were upright, inverted or negative. The statistics of the GLM fits are shown in Table 2. In all cases, the slope and intercept of the fitted logistic curves were significantly different from zero at $p < 2\text{e}{-16}$. The c-index, which measures the concordance between predicted and observed responses, was around 0.65 with 95% confidence intervals that did not overlap with 0.5 (the c-index value when there is only a random relation). Using simulations which assumed that the fitted logistic regressions generated the Bernoulli responses, we found 95% confidence intervals for the resulting c-index values. For upright, the 95% CI for the c-index is 0.65–0.68 (observed was 0.67); for inverted, 0.59–0.62 (observed was 0.60); for negative, 0.63–0.66 (observed was 0.65). Thus the c-index values were of the size expected if the model fit perfectly and the Bernoulli nature of the responses was the only cause of variability. These concordance values are slightly lower than those in Table 1. The reason is related to the slopes of the fitted regressions, which are shallower here. The situation yielding the highest c-index is one where x-values greater than some cut-off always produce a 0 response, and those below the cut-off always produce a 1 response. This will happen with a psychometric function that is a step. When the psychometric function is shallower, the predictor is no longer perfect because some 0 s occur above any cut-off and some 1 s occur below it.

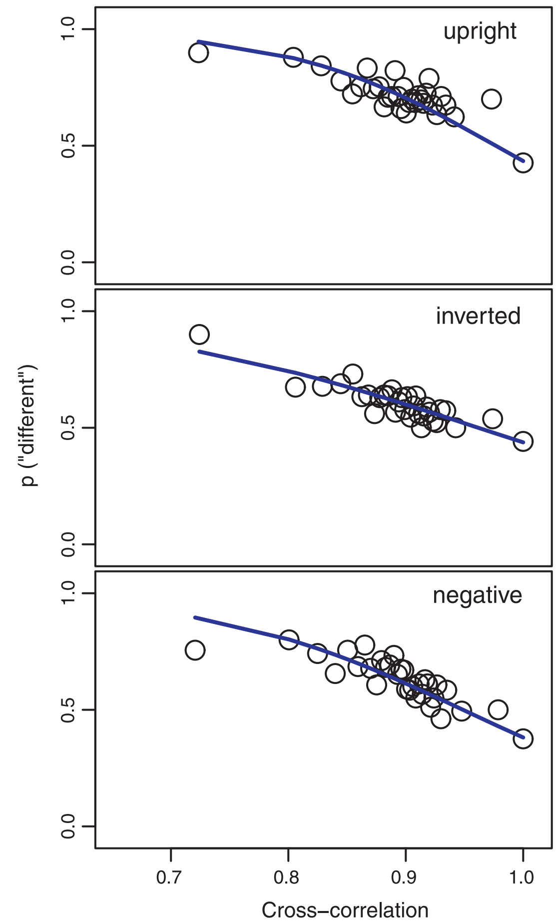

Fig. 5. The proportion "different" judgements as a function of the cross-correlation between the face pairs. Faces were upright, inverted, or contrast-reversed (negative). The plotted points are the proportions of responses in cross-correlation bins whose widths were adjusted to catch equal numbers of observations. Generalized linear model (logistic) fits to the unbinned 0/1 responses are shown.

The discrimination performance was best for the upright faces (slope = −11.34), followed by the negative (slope = −9.43) and lastly the inverted faces (slope = −6.58). Compared to the normal faces (upright and nonreversed contrast), the negative and inverted faces had significantly shallower slopes (upright vs negative: $z = 3.82$, $p = 0.00013$; upright vs inverted: $z = 9.64$, $p < 1\text{e}{-16}$). Thus face discrimination performance is impaired by inversion and contrast reversal, as has been found by other authors (Megreya & Burton, 2006; Riesenhuber et al., 2004; Vuong et al., 2005). The impairment can be modelled by an increase in the effective noise (as measured by the inverse slope of the psychometric function) added during computation of the cross-correlation. Studies by Gaspar, Bennett, and Sekuler (2008) and Sekuler et al. (2004) suggest that the main cause of higher effective noise with inverted and contrast-reversed faces is lowered sampling efficiency. Results from Sekuler et al. (2004) are consistent with observers using a template matching mechanism that is more efficient at using pixels around the eye and eyebrow regions when a face is upright rather than inverted. Inspection of their Fig. 2 shows that the loss of efficiency for inverted faces is due to a poorer template. This reflects the fact that observers have more expertise with upright (normal contrast) faces and thus will be able to form a better template for them. Both Gaspar et al. and Sekuler et al. conclude that performance with upright and inverted faces differs due to quantitatively rather than qualitatively different processing. Our results are in agreement with this conclusion.

Table 2
Statistics for logistic regression of p("different") as a function of cross-correlation between pairs of upright, inverted, and negative faces.

Condition	Intercept [95% CI]	Slope [95% CI]	χ^2	p	Concordance index [95% CI]
Upright	11.08 [10.10,12.05]	−11.34 [−12.36, −10.32]	548.60	<1e−16	0.67 [0.659,0.681]
Inverted	6.32 [5.46,7.19]	−6.58 [−7.49, −5.66]	210.22	<1e−16	0.60 [0.591,0.614]
Negative	8.95 [8.06,9.84]	−9.43 [−10.37, −8.49]	429.59	<1e−16	0.65 [0.642,0.664]

4. General discussion

Our results show that human discrimination of faces is well predicted by the cross-correlation between face images. The good prediction of performance by cross-correlation holds for faces that are frontal or profile, male or female, upright or inverted, and normal or reversed-contrast. Performance is impaired by face inversion and contrast reversal, showing that the discrimination task is capable of showing effects that are standard for face recognition. If we treat the observer as a black box, the output of the black box (proportion of "different" judgements) can be predicted on the basis of the cross-correlation between the input face pairs.

If the face discrimination performance can be predicted by low-level cross-correlation, does this imply that discrimination (as opposed to recognition) does not engage face-specific mechanisms? We found that performance was impaired by contrast reversal and inversion of the faces, as is normally found using a recognition task. Despite the impaired level of performance, it was still well predicted by the cross-correlation between the faces. Our results are consistent with those of Megreya and Burton (2006). Like us, they found that performance for inverted faces was impaired. They also found that discrimination performance for the upright and inverted faces was highly correlated, suggesting that similar processes (such as cross-correlation) may be involved in both. We have shown that the slope of the same/different discrimination function is lower when the face is inverted or contrast reversed. This raised level of effective noise could be due to more internal noise being introduced during the computation of cross-correlation when the faces are nonstandard, or to the sampling efficiency being lowered. As Sekuler et al. (2004) show, upright and inverted faces produce similar classification images. However the classification image for the upright images was somewhat better, producing a higher efficiency in using the information present. This conclusion about improved efficiency for normal faces was confirmed and extended to contrast reversed faces by Gaspar, Sekuler, and Bennett (2008). These issues of differences in efficiency must be addressed by any model of face discrimination.

In our experiments, the stimuli and task were kept as simple as possible. The judgement was whether the two images being presented were identical or not. So, for example, we never presented a frontal and a profile image of the same person, with the correct answer of "same". We also did not present face pairs with randomly varying size or orientation in the image plane. We did this in order to have the same sort of stimulus control used to study arbitrary patterns (e.g. Gabor patches). This does not make our experimental set-up completely artificial. For a very large proportion of real-world face discrimination scenarios, both by humans and machine (e.g. passport photo verification of identity), the faces being compared are viewed frontally. When a border guard examines a passport photo to verify personal identity, although the passport photo is small on the page, its visual angle is comparable to that of the real person standing in front of him or her, because the passport is held close and the person is distant. The guard holds the photo in a vertical orientation, in alignment with the person. Obviously any photo will not be identical to the person standing in front of the guard due to lighting, pose, expression and other factors, and indeed variations in these parameters will cause degradation in human performance.

One powerful benefit of using cross-correlation between faces as a way to predict performance is that it allows us to compare discrimination performance for types of stimuli that seem to be completely disjoint. Cross-correlation can serve as a common currency. So for example, one might compare discrimination performance for faces to that for gratings, measuring both as a function of the cross-correlation. Similarly, using cross-correlation would allow comparisons of performance for various face-based tasks, such as discrimination and recognition. Simpson, Falkenberg, & Manahilov, 2003, showed how cross-correlation allows performance for a variety of motion perception tasks to be compared.

In conclusion, we have shown that face discrimination performance can be predicted on the basis of the cross-correlation between the face images. Even if cross-correlation is not the mechanism used by the visual system, it can be pragmatically used by experimenters as a good predictor of performance. For more complex tasks, such as the situation where "same" faces could have different lighting or poses, a simple cross-correlation scheme would fail. To handle such cases, either the cross-correlation must be preceded by normalising operations, or a different mechanism altogether must be used.

References

Appelle, S. (1972). Perception and discrimination as function of stimulus orientation. *Psychological Bulletin, 78*, 266–278.

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology, 58*, 71–78.

Benton, A. L., & Van Allen, M. W. (1972). Prosopagnosia and facial discrimination. *Journal of the Neurological Sciences, 15*, 167–172.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London, 352*, 1203–1219.

Burgess, A. E. (1990). High level visual decision efficiencies. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 431–440). New York: Cambridge University Press.

Burgess, A., & Ghandeharian, H. (1984). Visual signal detection. I. Ability to use phase information. *Journal of the Optical Society of America A, 1*, 900–905.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*, 93–94.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3*, 1–8.

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 13988–13993.

Eckstein, M. P., & Ahumada, A. J. Jr., (2002). Classification images: A tool to analyse visual strategies. *Journal of Vision, 2*, i.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19*, 207–208.

Gaspar, C. M., Bennett, P. J., & Sekuler, A. B. (2008a). The effects of face inversion and contrast-reversal on efficiency and internal noise. *Vision Research, 48*, 1084–1095.

Gaspar, C. M., Sekuler, A. B., & Bennett, P. J. (2008b). Spatial frequency tuning of upright and inverted face identification. *Vision Research, 48*, 2817–2826.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *The Journal of the American Medical Association, 247*, 2543–2546.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Lehky, S. R. (2000). Fine discrimination of faces can be performed rapidly. *Journal of Cognitive Neuroscience, 12*, 848–855.

Lewis, J. P. (1995). Fast normalized cross-correlation. *Vision Interface*, 120–123.
Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research, 45*, 2287–2297.
Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). FMRI evidence for the neural representation of faces. *Nature Neuroscience, 8*, 1386–1390.
Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34*, 865–876.
Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London B, 271*, S448–S450.
Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*, 391–396.
Simpson, W. A., Falkenberg, H. K., & Manahilov, V. (2003). Sampling efficiency and internal noise for motion detection, discrimination, and summation. *Vision Research, 43*, 2125–2132.
Tanner, W. P., & Birdsall, T. G. (1958). Definition of d' and η as psychophysical measures. *Journal of the Acoustical Society of America, 30*, 922–928.
Troje, N., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*, 1761–1771.
van Trees, H. L. (1968). *Detection, estimation, and modulation theory*. New York: Wiley.
Vuong, Q. C., Peissig, J. J., Harrison, M. C., & Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Research, 45*, 1213–1223.
Whalen, A. D. (1971). *Detection of signals in noise*. New York: Academic Press.
Wilbraham, D. A., Christensen, J. C., Martinez, A. M., & Todd, J. T. (2008). Can low level image differences account for the ability of human observers to discriminate facial identity entity? *Journal of Vision, 8*, 1–12.
Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145.
Yovel, G., & Kanwisher, N. (2005). The neural basis of the face-inversion effect. *Current Biology, 15*, 2256–2262.